DOCUMENT RESUME

ED 072 115 TM 002 357

AUTHOR	Diederich, Paul B.
TITLE	II. Research 1960-1970 on Methods and Materials in Reading.
INSTITUTION	Educational Testing Service, Princeton, N.J.; ERIC Clearinghouse on Tests, Measurement, and Evaluation, Princeton, N.J.
SPONS AGENCY	National Inst. of Education (DHEW), Washington, D.C.
REPORT NO	ERIC-TM-22
PUB DATE	Jan 73
NOTE	10p.
EDRS PRICE	MF-$0.65 HC-$3.29
DESCRIPTORS	Initial Teaching Alphabet; *Phonics; Reading Achievement; Reading Development; *Reading Instruction; Reading Materials; *Reading Research; Reading Skills; Research Needs; *Research Reviews (Publications); *Resource Materials; Sight Method; Sight Vocabulary; Tables (Data); Teaching Methods; Textbook Evaluation; Vocabulary Development; Word Recognition

ABSTRACT

This bulletin summarizes and interprets some of the main findings of "Survey of the Literature on Methods and Materials in Reading," by Martha J. Maxwell and George Temp, Chapter IV of "The Information Base for Reading: A Critical Review of the Information Base for Current Assumptions Regarding the Status of Instruction and Achievement in Reading in the United States," the final report of a study prepared for the U. S. Office of Education (see ED 054 922). The bulletin focuses on the different methods used to teach reading and the differences in their results. Three previous summaries of research are evaluated, and the rationale for this study is given. Methods and criteria used to screen studies for review and criticism in "The Information Base for Reading" are presented, and the classification and ratings given the studies are provided. It is reported that little valuable information was extracted from this survey, due to the ineffectiveness of educational research in general. It is recommended that future research concentrate on the reasons children lose interest in reading after grade 8 and on ways of helping students acquire a large enough vocabulary to make the transition from juvenile to adult books. The adoption of whole-word methods, the Bliesmer-Yarborough Study, and the Initial Teaching Alphabet are reviewed. A bibliography is provided. (For related documents, See TM 002 358, 385) (KM)

ED 072115

ERIC CLEARINGHOUSE ON TESTS, MEASUREMENT, & EVALUATION
EDUCATIONAL TESTING SERVICE, PRINCETON, NEW JERSEY 08540

Conducted by Educational Testing Service in Association with Rutgers University Graduate School of Education

TM Report 22 January, 1973

TM 002 257

II. RESEARCH 1960-1970 ON METHODS AND MATERIALS IN READING

Paul B. Diederich

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
OFFICE OF EDUCATION
THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL OFFICE OF EDUCATION POSITION OR POLICY

This bulletin is a digest and interpretation of some of the main findings of Chapter IV, "Survey of the Literature on Methods and Materials in Reading," by Martha J. Maxwell and George Temp in *The Information Base for Reading: A Critical Review of the Information Base for Current Assumptions Regarding the Status of Instruction and Achievement in Reading in the United States*, the final report of a study directed by Reginald Corder (ETS, Berkeley Office) for the U.S. Office of Education, Project 0-9031, 1971. The full report is available in hard copy or microfiche through ERIC, ED 054 922.

The study was a survey of research from 1960 to 1970 bearing on three problems: the nature and extent of the current deficit in functional literacy, the effectiveness of different methods of teaching reading, and the training of teachers of reading. Using all possible bibliographic sources, the project staff listed over 15,000 documents bearing on these three problems. These were rated independently by five experts, and 1,855 were selected for critical review, including 741 on the second problem (the subject of this summary), 120 on the first and second, 56 on the second and third, and 31 on all three. Thus the chapter here summarized was based on a critical look at 948 documents bearing on methods and materials used in the teaching of reading. The list of all 1,855 documents that were reviewed occupies 134 pages of the final report.

The reviews were done by 22 doctoral candidates at the University of California in Berkeley. Applicants for this job all reviewed the same article, using a standard review form of eight pages developed by a technical committee. The most proficient participated in several training sessions and were monitored thereafter by the staff member who synthesized the reviews in each area. The reliability of those aspects of the reviews that could be quantified was determined by having 200 articles reviewed independently by two readers. The coefficients were all above .70 except one of .62 for a rating on "treatment," which was the most sketchily reported. This use of doctoral candidates as reviewers forestalled the objection that established researchers are hypercritical of the research of others. These young students were bent only on extracting whatever solid information they could find in the published reports.

Questions to Be Answered

This part of the survey was directed toward the following questions:

1. What methods, materials, approaches, equipment, and procedures are used to teach reading and to what extent?
2. What methods of reading instruction are built on essentially different pools of basic knowledge?
3. How much time and resources are expended directly on developmental and remedial reading instruction?
4. What relationships between methods of reading instruction and reading achievement of the various subgroups of the population can be shown?

Question 3 could not be answered from the literature reviewed. Even when it was important to hold instructional time constant, the difficulty of doing so was indicated by Stauffer (1967) in his introduction to the reports of 27 comparative studies supported by USOE:

> "Reading instruction time could not be defined acceptably.... Much effort was devoted to an attempt to define reading instructional time at the Coordinating Center meetings, but to no avail."

As for resources, the first guide to materials for reading instruction published by ERIC listed and described 10,000 items (Harris, L. A., 1968) and a supplementary list of 180 pages was issued the following year (Berridge, W. E. and Harris, L. A., 1969). No figures that would warrant even an estimate of the total sales of such materials were found.

Question 2, "What methods of reading instruction are built on essentially different pools of basic knowledge?" was dealt with conscientiously in a section headed *Knowledge Base* for each method reviewed, but the differences noted were too trivial to report in this brief summary. The impression given is that proponents of each method can find psychological or linguistic principles to support their views; that these principles had very little to do with the invention or development of their methods; that none of

This publication was prepared pursuant to a contract with the National Institute of Education, U. S. Department of Health, Education and Welfare. Contractors undertaking such projects under government sponsorship are encouraged to express freely their judgment in professional and technical matters. Points of view or opinions do not, therefore, represent official National Institute of Education position or policy.

these principles are affirmed by some schools of thought and denied by others; they are all commonly accepted principles; but proponents of each method seek out and emphasize those that support their views and ignore or underplay those that support competing methods.

In a few cases, it is hard to find any connection between the alleged basis for a method and the particular policies it advocates. The chief example is the work of Leonard Bloomfield, widely regarded as the father of modern linguistics, who also happened to be interested in methods of teaching beginning reading. He developed a set of exercises designed to teach children how the sounds of words they already knew were represented by letters. Instead of beginning with words of highest frequency in which the same letter represents different sounds, like *come, go, look, to* (from which it is hard to infer what the letter o stands for), Bloomfield restricted his initial vocabulary to words in which the letters stand for only one sound until that letter-sound correspondence is mastered, like "A fat cat ran at a bad rat." This restriction made it hard to write interesting stories, and his exercises appeared repulsively drill-like at a time when drill was at its lowest ebb of popularity and the newer "look-say" methods offered hope that reading could be learned with only gradual and incidental attention to letter-sound correspondences. It was twenty years before these exercises, revised by Barnhart, achieved publication as a result of Bloomfield's reputation as a linguist and Flesch's widely read attack on the neglet of phonics in *Why Johnny Can't Read* (1955). They were then described as "linguistically based" although it is hard to connect Bloomfield's revolution in linguistics with his commonsense preference for starting with words in which the letters represent only one sound until that sound is learned. He did, however, discuss reading and other practical applications in one chapter of his monumental work, *Language* (1933). Following his example, other initial reading materials have been described as "linguistically based" if the order in which new words are introduced is controlled, in part, by the ease and profitability of learning and the letter-sound correspondences that they exemplify.

Both Chall (1967) and the present survey also classify *English through Pictures* by I. A. Richards and Christine Gibson (1960) as "linguistically based," but in a totally different sense. Intended primarily to teach Basic English as a second language to adults, this series provides a gradual introduction to the structure of English sentences without resorting to grammatical terms, rules, or descriptions. The basic patterns of both word-order and word-endings are systematically exhibited in the minimal vocabulary of Basic English with stick-figures to clue the relationships expressed. Although the authors do not regard themselves as professional linguists, their attention to the structure of English certainly exemplifies another aspect of linguistics. It must not be inferred, however, that this series is thereby better designed to teach reading to native speakers at the age of six. That was not its purpose.

Since question 3 (time and resources expended directly on reading instruction) could not be answered at all from the literature examined in this survey, and question 2 (different pools of basic knowledge) yielded only trivial differences, we are left with question 1 (what the different methods are) and question 4 (differences in their results).

Previous Summaries of Research on Reading

The most comprehensive and accessible recent summaries of research on reading are those by William S. Gray (several articles following the general heading *Reading* in the 1960 *Encyclopedia of Educational Research*) and by David H. Russell and Henry R. Fea in N. L. Gage (ed.), *Handbook of Research on Teaching*, ch. 16, pp. 865-928 (Chicago: Rand McNally, 1963). These are both of the type needed by researchers who have to locate previous studies in an area they may want to investigate: a quickly comprehensible map of the domain, with studies in each area summarized in a sentence or two, most of them composed by graduate students. No one expects to learn anything from these brief statements; their purpose is only to locate studies that sound as though they might have a bearing on the question in which one is interested One can then find these studies in the bibliography and, if one is lucky, in an education library or on inter-library loan or through ERIC. It is only when these documents are read that information is transmitted—if any was there to start with.

A much more readable and informative type of summary is represented by Jeanne Chall, *Learning to Read: The Great Debate* (New York: McGraw-Hill, 1967). It is unified around a central hypothesis: that a preponderance of acceptable research from 1910 to 1965 favored early and systematic treatment of letter-sound relationships (phonics) despite the doctrine and practice prevailing from about 1930 to about 1960 that words should be learned as wholes, embedded in interesting stories, and only later, gradually, and incidentally analyzed into sounds represented by letter-combinations and pronunciation rules. The search for evidence pro and con has the fascination of a detective story, including the fictional detective's cleverness in showing that some of the leading investigators misinterpreted their own findings. Since a whole generation had been taught to read by the whole-word method, its faults rather than its achievements had become the focus of attention. A whole generation of parents had also become fed-up with Dick and Jane stories (which probably seemed more insipid and goody-goody to adults than to children), and their discontent was angrily expressed by Rudolph Flesch in his sensational best-seller, *Why Johnny Can't Read* (1955). Although Flesch oversimplified the problem and misinterpreted much of the research he reported, he touched off an explosion of popular feeling in favor of a return to phonics.

This controversy had such wide ramifications that it led Chall through a fairly comprehensive survey of research on reading from 1910 to 1965 and into a devastating analysis of the most widely used basal readers, including the voluminous manuals that tell teachers how to teach each lesson. Despite some criticism, this survey is accepted by most authorities on reading as careful, thorough, and sound in its main conclusions. It is almost a miracle that anyone could make such a mass of research on reading so readable. Since this book is now in almost every library, both educational and general, and since there is no present prospect of publishing the report here summarized, *The Information Base for Reading*, educators who wish to delve more deeply into research on reading will do well to start with Chall.

Since such an uncommonly good summary of research had recently been published, it may be hard to understand why the U.S. Office of Education commissioned another survey of research on reading from 1960 to 1970. First, the chapter here summarized on methods of teaching reading and their results is only one of three. The other two—the nature and extent of the current deficit in functional literacy and the training of teachers of reading (summarized in other ERIC bulletins)—were not touched by Chall. Second, the Office of Education was about to launch a multi-million dollar program of research on reading as part of the "Right to Read" program, designed to give every child by age 10 the basic skills that would later develop into the kinds of reading ability actually utilized by American adults in the conduct of their affairs—vocational, civic, and recreational. Such an all-out research effort required a closer look at a more comprehensive collection of recent research on reading than one could count on getting from Chall. She was admittedly and quite properly selective; she analyzed research over a long period of time bearing on her central hypothesis, concerned with the effects of early and systematic instruction in phonics. Although her search cut a wider swath through the literature on reading than anyone anticipated, there was no way to find out what was known for sure about reading, and what further research was needed, except to take a critical look at *all* serious research on reading, without regard to any central hypothesis, that had been reported in a limited period of time—the last decade.

The program of research then envisioned in the Office of Education was based to some extent on the "convergence technique" used by the National Cancer Institute of the National Institute of Health. It is a "systems approach" for the management of complex research programs when a large number of unknowns exist in the foundations of an area of science or technology. One of the first steps is to find out what has and what has not been demonstrated empirically in this area. For this reason, this survey of the literature was critical in a more general sense than Chall was critical. She reviewed a large number of the studies most often cited to find out whether the data warranted the conclusions reported by the investigator. The present survey started by applying certain canons of scientific research to all reports that passed the preliminary screening of an advisory committee. These canons were translated by Gephart (1970) into scales for profiling three dimensions of research quality: representativeness, treatment, and measurement. The points on these scales in abbreviated form were:

REPRESENTATIVENESS

5 Entire population
4 Random sample of specified population
3 Purposive sample of specified population
2 Volunteers
1 Unidentified group of subjects

TREATMENT

6 Theoretically based treatment described; controls for variables identified in the theory AND for extraneous variables that might have an effect
5 Same, but no controls for extraneous variables
4 Same, but no controls for either explicit or extraneous variables
3 No theory stated but treatment described in detail sufficient for replication
2 Commonly known treatment administered but not described in detail
1 Something of an unknown nature was done

MEASUREMENT

5 Standardized or ad hoc instrument; data presented which establish high validity and reliability for this application
4 Same with moderate validity and reliability
3 Standardized test but no data on validity or reliability for this application
2 Ad hoc instrument; evidence of moderate validity and reliability
1 Ad hoc instrument with no evidence of validity or reliability OR evidence of poor validity and reliability on either a standardized or an ad hoc instrument

A total rating of 8 or above out of a possible 16 on these three scales was classified as "acceptable," entitling the report to more thorough scrutiny than those classified as "unacceptable." In addition, the reviewers answered questions on the appropriateness of the data analysis procedures and whether the conclusions were supported by the data. It was recognized that these ratings could serve only as a rough screening device; that some studies with high ratings would yield little or no information on the questions under investigation, and that some with lower ratings would yield important information. But since the reviewers abstracted each article they read and answered many other questions about it, little information of value seemed likely to escape the notice of the staff member who synthesized the reviews in each area.

Classification and Ratings of Studies

Since the studies reviewed in this section of the survey were concerned with methods of teaching reading and their effects, the classification of methods used by Chall was adopted. The number of studies assigned to each category and the number and percent rated "acceptable" on the Gephart scales are indicated below.

These figures include some duplication when studies compared two or more methods. They do not include articles that could not be rated on the Gephart scales: theoretical, historical, descriptive, clinical, surveys, and reviews of the literature.

The number of acceptable studies at each level of schooling confirms Chall's finding that the bulk of research on reading has been directed toward the primary grades.

General Conclusions

The generally critical stance of this survey required an answer to the question: What important facts about methods of teaching reading and their results are incontestably known, so that further research may be directed elsewhere?

It must be stated at once that the answer yielded by this survey was "None." In view of the fact that more research has been done on reading than on any other subject taught in school, that more than a thousand studies are reported each year, that 265 of the studies reviewed were rated "acceptable" on the Gephart scales, and that everyone connected with this survey was motivated to locate findings that met basic standards of scientific research, this answer seems incredible. For one thing, what about the 265 studies that were rated "acceptable"?

A few of these studies will be examined in a later section to salvage whatever information they contain, but the general answer to this question is that, the more impeccable a study was (from the standpoint of research design and reporting), the more likely it was to reach the conclusion: "No significant difference in the results of different methods." The few significant differences that were found in acceptable studies by the end of grade 1 almost invariably disappeared by the end of grade 2 or 3. Slight but significant differences in one direction found by some acceptable studies were often balanced by slight but significant differences in the opposite direction found by other acceptable studies. While negative evidence can be valuable, the fact that no identifiable factors in the methods or materials of reading instruction appeared to make any consistent difference in properly designed studies can hardly be counted as valuable information on how to teach reading.

Although certain exceptions to this pessimistic conclusion will be noted in later sections, it may be advisable at this point to quote the actual conclusions reported by the investigators who directed this survey:

> No solid evidence or web of partial evidence exists at the present time that could restrict the teaching of a reasonably informed alert teacher of reading at any level. There are certain seemingly logical approaches and bits

Table 1. Classification and Ratings of Studies of Methods in Reading

METHOD	Number read and rated	Number rated acceptable	Percent rated acceptable
1. Meaning emphasis	122	63	52%
2. Code emphasis	50	31	62%
a. Synthetic (start with letters)	14	6	43%
b. Analytic (start with words)	10	3	30%
3. Linguistics	47	17	36%
4. Modified alphabet	40	21	52%
5. Responsive environment	7	3	43%
6. Programmed instruction	51	15	30%
7. Individualized reading	99	32	32%
8. Language experience	62	22	35%
9. Eclectic or combination	110	52	47%
TOTALS	612	265	43%

Table 2. Number of Acceptable Studies of Each Level of Schooling

Grades	1-3	4-6	7-9	10-12	College	Adult
Number rated acceptable	161	38	25	12	10	5

and pieces of information that might aid in deciding between certain aspects of procedures proposed, but literature exists both for and against almost any "method" that has been actually used or proposed in the last decade. If one has to take a position (and teachers obviously do every day of school), it is safe to say that:

> All methods of reading instruction instruct some children (probably the same ones) well and do not succeed with some small proportion of others that have been studied.
>
> The national reading problem is not that massive numbers of students cannot read in the sense of not knowing a graphemephoneme correspondences but that many persons do not wish to read for pleasure or information and do not comprehend either written or oral messages well.
>
> In effect, the national reading problem might just as easily be called the national thinking or comprehension problem, and the schools are only minutely responsible for the fact that massive numbers of our citizens are, essentially, not inclined to develop or maintain reading and comprehension skills necessary for their own self-selected goals and life space.

It may appear to many that the review of over 900 articles on reading methods and materials...ought to yield more information. As a matter of fact, it has been a constant concern that such an extensive, time-consuming, and detailed effort should yield only a little information. What standards should be applied was often asked.

However, to take seriously the charge to take a critical look at the literature relevant to the important questions of this study was to be unable to reduce the standards. Criticism is, unfortunately, judgment according to a standard. If the standards were too high, it was because they are the accepted standards of conducting and reporting research that science has found necessary over the years in utilizing the reports of its practitioners. These standards actually are flexible and vary from field to field. Now, however, is the time to refocus the research effort in reading.

> It is time to raise the standards of reporting and conducting reading research because it is impossible to know what trust to put in an incomplete research report, a study done with no controls, or a statement of conclusions unsupported with reported and clearly interpretable data. (pp. 136-137)

This pessimistic view of the bulk of reading research is supported by others who have taken a close look at studies in this field. Compare what Chall says of the literature she examined:

> Most (studies) did not indicate how the experimental and control groups were selected, how much time was allotted to various aspects of reading, how the teachers were selected, whether the quality of the teaching was comparable in both groups, or even whether the teachers followed the methods under study. Even more important, most studies did not specify clearly what a "method" involved, but instead merely assigned labels (e.g., "phonics"), expecting the reader to understand what was meant. (pp. 100-101)

Similar criticisms were expressed by Stauffer (1967) in his introduction to the reports of 27 first grade reading studies in a cooperative project funded by the U.S. Office of Education in the middle sixties:

> What conclusions can and cannot be made about the twenty-seven studies reported herein? Because of the many variables that were not uniformly controlled in the separate studies, the studies should not be compared to see which ones were most productive. Wide variation was found among the teachers involved....
>
> No one method should be compared with another because the methods were not sharply and clearly different. For example, all the methods included instruction in phonics of one kind or another, the alphabet, writing experiences, comprehension, and so on. Methods that were given the same label were not always the same....
>
> Reading instruction time could not be defined acceptably.... The tests used to measure readiness, intelligence, and achievement were not adequate.... In short, any attempt to compare method with method or study with study could produce gross misunderstandings and false conclusions.

Thus it is evident that the failure to extract important information from the literature examined in this survey cannot be ascribed to the jaundiced views of those who conducted the investigation. It is simply a fact that educational research in general and methods studies in particular have not yet achieved anything like the rigor or results of research in such fields as physics, agriculture, or genetics.

Probabilities, Possibilities, and Unknowns

Although one has to accept the conclusion that nothing is known for certain about methods of teaching reading, it would be unfortunate if the Office of Education decided simply to replicate all existing research with stricter controls and more adequate reporting. It would be doubly unfortunate if the present over-emphasis on the initial stages of learning to read were continued, coupled with neglect of later stages in which our failure to attain reasonable goals is much more apparent. With allowance for the percentage of failure that seems unavoidable in all human enterprises, it seems reasonably safe to say that we *do know* how to get practically all children past the initial stages of learning to read. Above grade 3, in fact, traces of any initial

advantage of one method over another are hardly ever found. Furthermore, except for severely deprived children, we are able to get most of them interested in reading books on their own—especially the series of adventure stories known in the book trade as "juveniles." In spite of the competition now offered by television, this interest in reading books apparently reaches a peak in or near grade 8 and then declines until the majority of American adults read hardly any books except the Bible (as a devotional exercise) and books involved in their work.

This last point is made with some misgivings, since it is one of the neglected areas of research on reading, but one bit of evidence is that, in a nationwide Gallup poll reported by TIME in July 1965, 77 percent of the adults questioned said they had not read a book during the previous year, while of a comparable sample in West Germany, only 33 percent said they had not read a book. In a recent survey of one day's reading of over 5,000 adults (another project of the "Right to Read" program), no directly comparable evidence could be found, since the respondents were asked only to review their reading of the 24 hours preceding the interview, but the result was at least compatible with the Gallup poll: only 33 percent of the sample reported reading anything in a book. If it is true, as Waples (1940) found in the thirties and the writer of this summary in the sixties, that this decline of interest in book-reading sets in at or near the onset of puberty, the causes are unknown, but it is a plausible hypothesis that this is the point at which the less capable readers finally have to make a transition to adult books if their reading interests are to continue.

There are countless reasons why this transition may be difficult, but one that is so obvious that they themselves and their teachers are aware of it is the harder vocabulary of adult books. There is a real dichotomy here: both authors and editors of juveniles constantly substitute easier words and expressions for those that young readers may not understand, but adult books—even detective stories—make no such concessions. If a student has not developed a recognition vocabulary adequate for adult reading, he is out of luck.

This brings us back to a point about early reading that was ignored in the literature examined in this survey. The publishers of nearly all basal readers somehow hit upon the standard of introducing not more than two new words per hundred running words, and as Chall demonstrates, they very rarely exceed this standard up to the end of grade 3; how far beyond they stick to it is unknown. These words are not new in the sense that the pupils do not know what they mean; they are familiar words that have not previously been used in that series. They may be recognizable on the simplest phonic principles; if the pupils know *sing*, they should have little difficulty with *ring* or *wing*, but they count as new words just the same.

Furthermore, it has become standard practice for the teacher's manual to tell which words are new in each lesson and to suggest ways of teaching the new words before pupils read the story in which they are introduced. For example, if the only new words are *pretty* and *too*, the teacher is advised to choose a pupil's painting and write under it, "This is pretty," and under another, "This is pretty, too." These are displayed with enough discussion to make the meaning obvious, and pupils are asked to read these sentences and show their understanding in some such way as pointing to something else that they think is pretty. Then they are ready to read the story in which these are the only new words.

With this amount of build-up required for each lesson, it is understandable that primary teachers would object to a much heavier vocabulary load than two new words in a hundred. It is also obvious that when pupils come to these words in the story, there is no need to apply any phonic principles to figure out what sounds these letters represent, for they have already been taught to recognize these words at sight. Finally, when they come to the fourth part of the typical lesson that is supposed to be devoted to word-analysis to discover and drill upon letter-sound relationships and pronunciation rules—the hardest and most abstract part of the lesson—there is likely to be little time or energy left for it and certainly no motivation, for the pupils have already learned the words that the rules might enable them to decode.

Thus the light vocabulary load characteristic of basal readers and the accepted method of teaching the new words seem likely to induce in pupils and teachers alike a mind-set that will make it impossible to acquire a recognition vocabulary adequate for adult reading by age 13. Where did this absurd standard come from? Chall, who has an article on "The History of Controlled Vocabulary" (1958), does not know, and no basis for it was found in any of the literature examined in this survey. One possibility is that Michael West, who had a successful series of readers for teaching English as a second language in India, visited this country during the thirties and presented some impressive evidence that a vocabulary load of two new words per hundred was about right for his purpose. Of course, he was talking about words that were completely unfamiliar to Indian students—not words that they used every day of their lives but had not previously seen in print. There was also the very important point that after West worked up to a vocabulary of about 1,500 words, he stopped counting, confident that any new words introduced thereafter would be surrounded by enough familiar words to insure a tolerable rate of acquisition. West never intended his findings to be applied to initial reading materials for native speakers at the age of six, but since the basal readers embodying the new "whole word" approach were then *in utero*, it is possible that they were influenced, consciously or unconsciously, by West's findings.

Whatever the source may have been, the standard of two new words per hundred and the practice of teaching the new words before allowing pupils to read the stories in which they occur should both be subjected to experimental

verification. So far as we can discover in this or any other recent survey of the literature, this has never been done, and any experimental test would probably expose both the standard and the practice as unfounded and pernicious. They make it virtually impossible for any but the top quarter in verbal ability to develop an adequate reading vocabulary by age 13, and the absence of such a vocabulary by that age is the most probable single cause of the decline of interest in book-reading thereafter. If we are not as great a nation of book-readers as Russia or Japan—a question that seriously affects our national interests—the root of the problem may lie right here, but it would continue to be overlooked by any replication of existing research.

The Achievement of the Whole-Word Methods

At this point in time it seems obvious that the proponents of the whole-word method were wrong in trying to teach nearly all previously unused words as though they were Chinese characters, with only later and incidental attention to the sounds represented by letters. In the next section we shall try to demonstrate that one of the few conclusions of reading research in which we can have a high degree of confidence is that earlier and more systematic instruction in phonics is essential. But we must not forget the situation that the whole-word method was designed to correct and the excitement generated by its early successes, or we shall swing back with the pendulum to the same instructional errors that were perpetrated in the twenties.

When the writer of this summary began his graduate study of education in 1928, he was told by no less an authority than Walter Dearborn that it was quite possible to learn to read without first learning the alphabet. This seemed preposterous, and Dearborn had to send his students to observe several classes that were learning to read by the new "look-say" method before they would believe that it was possible. Then they were told to read some of the early research of the men who were to dominate the field of basal readers for the next thirty years, Arthur Gates and William S. Gray, and it began to seem not only possible but also advantageous in some respects to begin reading instruction without assuming or first developing knowledge of the alphabet.

What the current innovator tends to forget is that initial instruction in letter-sound relationships and pronunciation rules was done to death at that time. The children had to learn so much abstract material by rote before doing any significant amount of reading that the situation was the opposite of the one we face now. Children now have little reason to learn the rules for decoding because they have already been taught the words that the rules might have enabled them to recognize. In the twenties, children had even less reason for learning these rules because they had not yet encountered these words in any story that held their interest, but only in the few words that were chosen as examples. The result was that the early stages of reading instruction were a titanic struggle between the teacher and her class, and many more children fell by the wayside than do so at present.

When prospective teachers like the students of Walter Dearborn discovered what a relatively painless process the teaching of reading could be, using the new "look-say" or whole-word approach, they were not disposed to demand evidence of superior results by the end of the year. It was enough to show that the new method worked about as well as the old and with far less agony. That is why so many studies reviewed by Chall and even some included in this survey simply compared the "meaning emphasis" with the older "code emphasis," and why the investigators were satisfied with the finding that one worked about as well as the other. It took the sharp eyes of Jeanne Chall to detect the fact that a stronger "code emphasis" than the look-say method would permit in the early stages actually held a slight edge throughout, and when the kind and amount of phonics instruction were adjusted to what children could stand, a very decisive advantage. Now that this finding is accepted, we must be on guard against the fallacy that, if a moderate amount of phonics from the start is advantageous, a large amount will be still better. That would bring us right back to the twenties and set the stage for another revolution with renewed emphasis on meaning.

The Bliesmer-Yarborough Study

The case for early and systematic phonics instruction is epitomized by a "landmark" study by Emery P. Bliesmer and Betty H. Yarborough, "A Comparison of Ten Different Beginning Reading Programs in First Grade," *Phi Delta Kappan*, 1965, *46*, pp. 500-504. It was published after Chall's chapter summarizing the experimental evidence had been written and therefore received only postscript attention, but it sharpened the conclusions she had teased out of the previous fifty years of research. It is the only individual study that is analyzed in detail in the full report of the chapter here summarized, and this analysis included some technical criticism that would be out of place in a brief summary. Only the methods and main findings will be presented here in order to illustrate one of the few conclusions of research on reading that have been established beyond reasonable doubt.

Bliesmer, director of the McGuffey Reading Clinic at the University of Virginia, selected five reading programs representing a "synthetic" approach to phonics instruction: early and systematic attention to letter-sound relationships and pronunciation rules and using them to "synthesize" whole words (i.e., sound them out). He also selected five programs representing an "analytic" approach: starting with a large number of whole words that children were taught to recognize at sight and later analyzing these words to discover and drill upon the sounds these letters repre-

sent. It must be understood that both groups of programs undertook to teach phonics, but one group did it early and systematically, the other later and incidentally.

He then persuaded Mrs. Yarborough, director of developmental reading for the Chesapeake (Va.) public schools, to try out all 10 programs under rigorous experimental conditions in four elementary schools, two middle-class and two slightly lower but not lower-class. Each program was tried out in two schools, one middle-class, one lower, by one teacher in each school. With this restriction, the principals drew lots as to which programs would be tried out in their schools, and teachers were randomly assigned to these programs except that no inexperienced teacher was eligible and no teacher rated less than "average" in teaching skill during the previous year. These teachers were given intensive training in the program they were to teach and an unusual amount of supervision throughout the year, not only by the language arts supervisors but also by publishers' representatives. It was agreed that no pupil should receive more than 45 minutes of reading instruction per day, and the supervisors reported that this limit was strictly observed.

All first graders in these four schools were randomly assigned to classes so that neither the pupils nor the teachers had any choice of one another or of the reading program they were to follow. The Metropolitan Readiness Test and the California Mental Maturity, Short Form, were administered at the start and the Stanford Achievement Test, Primary I, Form W, at the end. The latter yields five part-scores: word reading, paragraph meaning, vocabulary, spelling, and word-study skills. Complete data were obtained on 484 pupils: 236 in the synthetic programs (systematic phonics), 248 in the analytic; 112 were lost through moving or incomplete data. Analysis of variance revealed significant differences among the randomly constituted groups in age, verbal and nonverbal IQ, and total readiness. These data were used to "adjust" the average scores for each program on the final reading test. The brief report in *Phi Delta Kappan* does not indicate how this was done except by saying "analysis of covariance." A statistician would like to know more than this, but this is the accepted technique for controlling initial differences that remain after randomization, and it is hard to think of any plausible way in which it could introduce any systematic bias in favor of either group of programs. The Division of Educational Research, University of Virginia, assisted in analyzing the data.

In comparing five synthetic (systematic) with five analytic (incidental) phonics programs on five tested outcomes of the year's work, there were 125 comparisons, and in 93 of these there was a significant (.01) difference in favor of the synthetic programs and not one in the opposite direction! This was almost too good to be credible. It was not a case in which the differences were statistically but not educationally significant. The differences were expressed in grade-level scores in which 1 was the average distance between grades 1 and 2 in the norms on the final test. On this scale the average difference between synthetics and analytics in all 125 comparisons, significant or not, was .49; and if we take just the 93 significant differences, the average superiority of the synthetics was .58—slightly more than half a year of normal progress at this level of schooling.

In the subtest on paragraph meaning (on which many "analytic" programs reviewed by Chall maintained an advantage up to the end of grade 1), there was a significant (.01) difference in favor of the synthetics in 20 of the 25 comparisons, and the average magnitude of these significant differences was .37 (of one year). This was the smallest margin of superiority among the five subtests; the others were .38 in word reading, .61 in vocabulary, .69 in spelling, and .83 in word-study skills. None of the five synthetic programs was outstanding; four were nearly equal in their average superiority to the analytics; but one was weak—a new programmed textbook that has since been revised.

By way of contrast, when the five synthetics were compared with one another, there were 50 comparisons but only 17 significant differences, averaging .50; among the analytics there were only six, averaging .30. No one expected the results of each group of programs to be more uniform than this, and these differences within groups were as nothing compared with the enormous difference between the two groups: 93 significant differences out of 125 comparisons, all in the same direction, averaging about half a year's work.

Initial Teaching Alphabet

The only other method reviewed in this survey with anything like a comparable number of significant differences to its credit was the Initial Teaching Alphabet (ITA) devised by Sir James Pitman in England. It uses all the conventional letters except q and x but introduces 22 other characters to make up the 46 significantly different sounds (phonemes) of English. The new characters resemble the traditional letter-combinations to which they are most closely related in sound in order to facilitate the transition to TO (traditional orthography) which usually occurs at the end of grade 1 or the beginning of grade 2. Most of the new characters (graphemes) are formed by joining together two traditional letters which are taught as one letter and one sound: *game*, for example, becomes *gaem* with a juncture between the a and e to represent long a. The additional burden on the memory of learning 46 characters is offset by the fact that no capital letters are used in ITA; capitals are replaced by a slightly larger size of the lowercase letters.

When teachers are shown the sort of story in ITA that children can read by the end of the first grade, it looks strange at first, but with a little concentration they find that they can read it without much difficulty. Most of the words are spelled as usual, and those that are not have only certain parts changed in ways that are easy to decode. This explains why children have so little difficulty with the

switch in the opposite direction: from ITA to TO. Although the two are so nearly alike at the point of transition from one to the other, the advantage of ITA is that it is much easier to learn in the initial stages because each character represents only one sound. First graders quickly learn how to recognize the ITA printed form of most of the 4,000 words they use instead of learning these words one at a time by sight. Hence a much larger vocabulary is available from the start; more interesting stories can be written; children can start reading more quickly and read more than children struggling with TO. They can also write earlier, more easily, and more creatively, since there is relatively little trouble with spelling.

If one tries to exploit all these advantages at once, the purists will argue that there is no way to tell which of these features accounts for any superior results that may be found, but the promoters will argue that it is impossible to tell how superior the results might be until one combines all the advantages of a new medium, since they depend on and support one another.

Generally speaking, the English experimenters took one route and the Americans the other. In the first experimental trials in England, Downing (1969), an industrial psychologist, had the most popular English basal readers printed in ITA and told teachers to continue teaching them exactly as they had before except for the changes necessitated by teaching the new alphabet. After two years, the children who studied these readers in ITA and then switched to TO were about a year ahead of children who had studied the same readers in TO from the beginning.

The American experimenters, led by Mazurkiewicz (1967), refused to be tied down by the restrictions that had to be observed by readers designed to teach TO. They developed a new series in ITA with a heavier vocabulary load, earlier and heavier stress on letter-sounds, more reading, and more writing. Results were compared with those of conventional basal readers, and the differences were even more favorable to ITA than in the English experiments, especially after the transition to TO. The most comprehensive and impartial review of results of both English and American studies up to this point is by F. W. Warburton and Vera Southgate, *i.t.a.: An Independent Evaluation* (London: Newgate Press, 1969). We shall not quote actual figures in this case since results of the definitive experiments now in progress have not yet been published.

Concluding Comment

The survey and critical analysis of over 900 articles on methods and materials of reading instruction occupied 77 single-spaced pages, exclusive of bibliography, in the full report of this study. What can be reported in 15 percent of this space obviously represents just a few highlights. We made no attempt to say something about every part of the survey, believing that full treatment of the findings that seemed most interesting and important would have greater impact. Omission from this digest does not indicate that a method was slighted in the full report, but it does create a presumption that no outstanding triumphs were recorded. For example, the most startling fact reported about programmed material was that research was available on only 7 percent of the products marketed up to 1971 while some field testing was claimed for another 8 percent (p. 89 in the full report). Computer-assisted instruction is still in its infancy. No one has found a way to overcome the reading deficiencies of groups that have been held down by poverty, segregation, and contempt. Removal of these handicaps seems more likely to produce improvement in reading than any change in method.

BIBLIOGRAPHY

1. Berridge, Wayne E., & Harris, Larry A. Guide to materials for reading instruction, supplement 1. September 1969. 180 p. ERIC: ED 032 452.

2. Bliesmer, Emery P., & Yarborough, Betty H. A comparison of ten different beginning reading programs in first grade. *Phi Delta Kappan*, 1965, *46*, 500-504.

3. Bloomfield, Leonard. *Language*. New York: Holt, Rinehart & Winston, 1933.

4. Chall, Jeanne. *Learning to read: The great debate*. New York: McGraw-Hill, 1967.

5. Chall, Jeanne. The history of controlled vocabulary. In J. Allen Figure 1 (Ed.), *Reading for effective living*. International Reading Association convention proceedings, 1958, 3, 177-180.

6. Downing, John. Initial teaching alphabet: Results after six years. *Elementary School Journal*, 1969, *69*, 242-249.

7. Flesch, Rudolf. *Why Johnny can't read and what you can do about it*. New York: Harper & Brothers, 1955.

8. Gephart, William J. Application of the convergence technique to basic studies of the reading process. Final report. Project No. 8-0737. Grant No. OEC-0-8-080737-4335. United States Office of Education, National Center for Educational Research and Development. 1970.

9. Gray, William S. Articles on reading. In Chester W. Harris (Ed.), *Encyclopedia of Educational Research*. New York: Macmillan, 1960. Pp. 1086-1135.

10. Harris, Larry A. Guide to materials for reading instruction. Indiana University, Bloomington. 1968. 277p. ERIC: ED 019 528.

11. Mazurkiewicz, Albert J. The initial teaching alphabet in reading instruction, evaluation—demonstration project on the use of I.T.A. Comprehensive final report. Lehigh University, Bethlehem, Pa. Feb. 1967. 81p. ERIC: ED 013 173.

12. Richards, I. A., & Gibson, Christine. *English through pictures*. New York: Washington Square Press, 1960.

13. Russell, David H., & Fea, Henry R. Research on teaching reading. In N. L. Gage (Ed.), *Handbook of research on teaching*. Chicago: Rand McNally, 1963. Pp. 865-928.

14. Stauffer, Russell G. (Ed.) *The first grade reading studies: Findings of individual investigations*. Reprinted from *The Reading Teacher*. Newark, Del.: International Reading Association, 1967.

15. Waples, D., Berelson, B. R., & Bradshaw, F. R. *What reading does to people*. Chicago: University of Chicago Press, 1940.

16. Warburton, F. W., & Southgate, Vera. *i.t.a.: An independent evaluation*. London: Newgate Press, 1969.